David Randahl

Who knows what tomorrow will bring?

Four papers on the prediction of contentious politics

UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in Sal IX, Universitetshuset, Biskopsgatan 3, Uppsala, Friday, 10 June 2022 at 10:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Bear Braumoeller (Ohio State University, Department of Political Science).

Abstract
Randahl, D. 2022. Who knows what tomorrow will bring? Four papers on the prediction of contentious politics. *Report / Department of Peace and Conflict Research* 128. 35 pp. : Uppsala University. ISBN 978-91-506-2948-4.

In the last decade advances in statistics, computing power, and data collection has led to an increased interest in forecasting within the field of peace and conflict research and to the adoption of a wide range of methodological approaches for making such forecasts. By making use of these more powerful forecasting methods researchers have been able to produce accurate predictions, as well as better inferences, of many different types of contentious politics events and to create operational early warning systems for such events. Adapting these forecasting methods to the social world in which politics and political behavior operate, however, is not without its challenges. This dissertation explores a number of methodological issues and advances in peace and conflict research, both inferential and forecasting oriented, through a series of four papers. In the first paper, I explore trends in democratization and autocratization using dynamic simulation. In Paper II, my co-author and I take aim at the difficulty of modeling and making forecasts with data which contains both excess zeroes and extreme-values. We propose an extreme-value and zero-inflated regression model which we use to replicate a study on the effects of UN peacekeepers on violence against civilians. Paper III explores latent variable modeling by using Markov models to make forecasts for escalation and de-escalation of armed conflicts. In the last paper, I investigate the effects of missing data and imputation techniques on the predictive performance of models. The four papers of the dissertation make several contributions to the growing literature of forecasting within peace and conflict research. First, the dissertation contributes to the methodological aspects of conflict forecasting by developing new statistical tools, Paper II, and adapting tools from other fields to different processes of armed conflict and contentious politics, Papers I & III, as well as by evaluating the practical effects of common choices in data pre-processing on the performance of forecasts in Paper IV. Second, the dissertation contributes to new ways of drawing inferences about conflict processes by anchoring the inferences in the latent state of the conflict processes in Papers II & III, and through the comparison of aggregated simulations to the historical record in Paper I. Lastly, the dissertation makes a substantive contribution to the broader field of peace and conflict research in Papers I & II by contributing to the debate on the waves of democratization and autocratization, and by nuancing the impact of UN Peacekeepers on violence against civilians.

David Randahl, Department of Peace and Conflict Research, Box 514, Uppsala University, SE-75120 Uppsala, Sweden.

ISSN 0566-8808
ISBN 978-91-506-2948-4
URN urn:nbn:se:uu:diva-473030 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-473030)

To those who made the tools which
others used to build great things...

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Inexorable Force or Dying wave? Long term trends of democratization and the third wave of autocratization

II Inference with extremes: Accounting for Extreme Values in Count Regression Models

III Predicting escalating and de-escalating violence in Africa using Markov Models. Published in *International Interactions* 48 (4) (2022, forthcoming).

IV What's missing? The effect of missing data and imputation techniques on predictive performance in forecasting civil war violence

Reprints were made with permission from the publishers.

Contents

Acknowledgements

This dissertation is in many ways the product of a lifetime of inquiry and curiosity. I guess few people would be surprised that the little boy who always asked questions, made (probably quite annoying) inductions about the world, and who wanted to become an inventor or a scientist ended up doing a PhD. The journey has, however, been a long one and there are many people I would like to thank, without whom this dissertation would not have been possible.

First and foremost I would like to thank my parents for providing me with a safe and loving home, for always supporting me in whatever endeavours I've pursued, and for providing a safe haven for me to return to whenever I've felt I needed a break and to just re-charge my batteries. I would not have been able to do this without your support. I would also like to thank my brothers for always being on my side, being there to listen, and for being great role models all around. Among my greatest inspirations in life have been my grandparents, who showed me that the norms set up by society do not define who we become. In a time where university education was still rare for girls, my grandmother became one of the few dentists on our beloved Gotland, and in a time where men were expected to be breadwinners of the family my grandfather did not at all mind that my grandmother was the breadwinner of theirs. In a way, I believe that their willingness to defy the norms of their time to pursue their own destinies and their own happiness has defined our family for generations.

Academically I was for a long while unsure which of all the potential topics in the world I wanted to pursue, until I met my main supervisor and academic inspiration Håvard Hegre. When Håvard introduced me to the idea of ViEWS I was immediately hooked, as it represented a combination of topics that I had grown more and more interested in, in the intersection of statistics and peace and conflict research. On that note I would also like to thank my three supervisors. Håvard, for always posing challenging questions and constantly inspiring new research ideas, finding the core of arguments and giving feedback on how to strengthen them, and always pushing for excellence. Espen, for always being available to listen and for helping me navigate the world of the PhD life, and for helping me prioritize and seeing the struggles outside of academia. Måns, for providing me with feedback on the statistical details and rigour needed to allow people outside my narrow field of vision understand what I've been doing. The members of the supervision group, the 'Monday group on Wednesdays', have also given me invaluable feedback on my project. In no particular order, I would like to thank Lisa, Hanne, Nina, Kristina, Mihai, Gulla, Ida, Eric, Sophia, Karin, Jenniina, Stefan, Marcellina, and Maxine, for the comments throughout the last five years.

Apart from my supervisors and supervision group, this dissertation would not have been possible without my phenomenal co-author Johan Vegelius, who have shown me what interdisciplinary is all about and has given me a blueprint for how to make successful co-authorships in the future. I hope we will continue to create plenty more fantastic research ideas over beers well into the future!

The Department of Peace and Conflict research is truly a unique place to work at. Not only have I been able to spend every working day with truly brilliant colleagues at the top of their respective fields, I've also built friendships with the most fantastic group of people (nerds) ever around. I'm privileged to have been working at such an amazing place where an after-work with your colleagues is the same thing as going out for beers with your friends. Without these beers, I would surely not have been able to finish this dissertation.

Believe it or not, I have also had a life outside the department and there I would like to especially thank Lovisa for being an amazing friend, colleague, and all around partner in crime, Ellen for being one of my best friends since always, Holly Langham and Kate (who's of course also in the department) for being the best support there is, Holly Guthrey for always being up for a trip to flavortown, and M, G & the Bs for all the silliness. All members of the *secret* group of ethanol imbued democrats, Simon, Oskar, Peter, Henrik, the Antons and the Axels, Viktor, and Jonatan, also deserve a special mention for providing continuity and tradition in an ever changing world. On a more serious note, I would also like to thank my therapist Barbara, who has given me so many more perspectives on my life and without whom this journey would probably have been much longer and much more painful, and Petra, for mutual support when things have been psychologically tough.

Last, but very much the opposite of least, Ronja. You bring such joy into my life every day, you are my rock and my favorite person all around. There is never a boring day when you are around, and I'm amazed and so incredibly grateful for being allowed to spend every day with you. I love you from the bottom of my heart, and I'm so very much looking forward to the rest of our lives together!

David Randahl
Uppsala, April 2022

Introduction

> *No matter how I turn it over in my mind, the number one task for peace research always turns out to be that of prediction: the ability to forecast, with increasing reliability, the outcomes which are most likely to emerge out of a given set of background conditions and behavioral events.*
>
> J. David Singer, 1973

Almost fifty years have passed since David Singer in 1973 identified prediction[1] as one of peace researchers' primary goals. Despite initial optimism and early attempts of forecasting in the years following Singer's ([1973] 2012) essay on foreign policy prediction (for instance Azar, 1980; Singer and Small, 1982), systematic and structural forecasting of conflicts with any practically useful degree of accuracy remained mostly out of reach for peace researchers in the following decades, due to a combination of insufficiently detailed data, limited methodological options, and the lack of the computational power required to use more advanced statistical forecasting techniques (Hegre et al., 2017; Schrodt, Yonamine, and Bagozzi, 2013; Beck, King, and Zeng, 2000). As the field of peace research increasingly moved towards explanatory studies focusing on understanding the drivers of different types of conflict behavior, both the feasibility of producing accurate forecasts (Stevens, 2012) and, out of the risk of producing self-fulfilling prophecies, the desirability of such forecasts was questioned (Collier, 2008).

In the last decade, however, the interest in forecasting among peace researchers has been rekindled. This renewed interest in forecasting is partly driven by methodological and computational advances, and improvements in data collection and quality, which has allowed more ambitious forecasting of conflict events and other forms of contentious politics (see for instance Hegre et al., 2013; Hegre et al., 2019; Gleditsch and Ward, 2013; Goldsmith et al., 2013; Chenoweth and Ulfelder, 2017; Beger, Dorff, and Ward, 2014), and partly by a criticism against the deficiencies of explanatory studies based on null-hypothesis significance testing (NHST) (see in particular Ward, Greenhill, and Bakke, 2010, and Schrodt, 2014) .

This dissertation is firmly set in this new generation of conflict forecasting and explores a number of methodological challenges and puzzles which arise when adapting forecasting models and research methods from other scientific

[1] Prediction and forecasting is used interchangeably throughout this introduction.

fields to that of peace research, and how to obtain better and more meaningful inferences in the face of these methodological challenges. An increased understanding of these challenges is fundamental in order to further advance research within quantitative, systematic, forecasting of different types of phenomena of interest to peace researchers. The four composite papers of this dissertation approach different aspects of the methodological and inferential challenges at the research frontier for forecasting in peace research. The dissertation takes an interdisciplinary approach in the borderlands of traditional peace and conflict research and statistics in order to approach these challenges. A common theme across the papers is that they are all either based based on methodology or data used in the Violence Early Warning System (ViEWS) (Hegre et al., 2019; 2021) or tackle methodological and forecasting problems relevant to the ViEWS project.

In Paper I, I use dynamic simulation to explore the trends in democratization and autocratization from 1789 to 2020 in order to evaluate the hypothesis that democratization and autocratization tend to occur in waves and that the 'third wave of democratization' has come to an end. This paper showcases how dynamic simulation, developed as a forecasting tool in ViEWS, can be used for theory evaluation and to obtain unique inferences which cannot be modelled using NHST. In addition, the paper contributes to the debate on the global trends in democratization and autocratization.

Paper II tackles the issue of extreme values in conflict research by proposing a new regression model, the extreme value and zero inflated (EVZINB) model, which better handles extreme values and offers new insights into the drivers of extreme observations. Extreme values are a particular methodological challenge for both inferential research and for forecasting purposes. In inferential research, extreme values tend to have an out-sized effect on the results, leading to possibly biased inferences and unstable estimation. In forecasting, extreme values pose a double challenge as their rarity make them difficult to forecast while they often are the cases most important to make accurate forecasts for. The EVZINB model offers a modelling framework where observations are conditional on a latent, unobserved, state which allows effects to be modelled separately for the different latent states. This allows for more stable estimation as well as novel inferences and better out of sample predictive performance compared to earlier approaches. The latent states can also be used directly in conflict forecasting to provide forecasts for the likelihood that a conflict enters into the 'extreme value latent state'.

Paper III continues exploring latent variable modeling by employing Markov models to forecast escalation and de-escalation of armed conflicts in Africa within the ViEWS 2020 prediction competition (Hegre, Vesco, and Colaresi, 2022). Conflict processes are known to have a high degree of spatio-temporal dependency which poses a challenge for forecasting far into the future. The proposed Markov modeling framework utilizes this dependency by making forecasts conditional on observed or hidden *Markov states*. The Markov states can be though of as the current dynamics of the conflict and can therefore be used both to obtain accurate forecasts, and to make these forecasts intuitive and explainable. The Markov models developed in this paper have also been shown to be powerful forecasting models in the ongoing ViEWS project for making forecasts of the level of fatalities across conflicts (Hegre et al., 2022) .

In Paper IV, I investigate the effects of missing data and data imputation techniques on the predictive performance of forecasting models. Missing data is a problem which plagues many disciplines of science and which is common in data sources used within peace research. Using the ViEWS fatalities prediction project (Hegre et al., 2022) as an empirical test case, I show what the effects of different types of missing data are on the predictive performance of different models, and how this effect is affected by different imputation techniques. I also show that the guidelines for how to handle missing data in inferential studies are not necessarily useful to guide handling of missing data in forecasting studies.

In all, the four papers of the dissertation make several important contributions to a growing literature of forecasting oriented studies within peace research. First, the dissertation contributes to the methodological aspects of conflict forecasting by developing new statistical tools, Paper II, and adapting existing tools from other fields to different processes of armed conflict and contentious politics, Papers I & III, as well as by evaluating the practical effects of common choices in data pre-processing on the performance of forecasts in Paper IV. Second, the dissertation contributes to new ways of drawing inferences about conflict processes by anchoring the inferences in the latent state of the conflict processes in Papers II & III, and through the comparison of aggregated simulations to the true historical record in Paper I. Lastly, the dissertation also makes a substantive contribution to the broader field of peace and conflict research in Papers I & II by contributing to the debate on the waves of democratization and autocratization.

This introduction to the dissertation aims to outline some of the methodological and theoretical advances and debates which underpin the current generation of forecasting within peace research and which form the foundation on which this dissertation stand, and to give a comprehensive summary to the four composite papers of the dissertation and their contribution to field. In the next section, I present a brief history of forecasting within the field of peace research. This is followed by a section where I outline some of the differences between predictive and explanatory modelling in peace research, as well as critiques against both frameworks, and how predictive and explanatory methods can be combined to generate and evaluate theory and produce more understandable and practically useful forecasts. I then summarize the four papers of the dissertation, describing in particular the research gap(s) the papers are motivated by as well as the findings and contributions of each paper. Finally, I discuss the wider implications of this dissertation for forecasting in peace research and suggest avenues for further research.

Systematic forecasting in peace research

Forecasting of conflict events and other forms of contentious politics has been on the research agenda for over fifty years. In the 1960s, data collection for conflict data was pioneered by Richardson (1960b) and the Correlates of Wars project (Singer and Small, 1982; Singer, 1972). This together with the adoption of statistical methods for the analysis of wars allowed researchers to systematically investigate the drivers of wars and armed conflicts and to calculate probabilities of wars and conflicts breaking out based on measurable national and dyadic indicators. While early quantitative peace researchers expressed skepticism of the ability of using statistical models for precise prediction of when and where wars would appear (Wright, 1964; Richardson, 1960a; Moses, 1961), these advances in data collection and methodology spurred a new generation of researchers focused on conflict prediction. The following decades saw further advances in data collection and several ambitious attempts at systematic forecasting of conflict and the creation of early-warning systems for conflicts and other types of crises (for instance Singer and Wallace, 1979; Andriole and Young, 1977; Azar, 1980) and even to the integration of statistical crisis forecasting in the daily brief for the US president (Kennedy, 2015).

The initial forecasting optimism faded in the early 1980s as it became increasingly clear that structural forecasting could not live up to the high expectations from researchers and practitioners (Schrodt, 1988). Instead, forecasting of foreign policy, wars, and crises, moved primarily to game theoretical expected utility models pioneered by Bueno de Mesquita's (eg. 1980; 1981; 1984; 1988) which was extensively used by the US government during the cold-war and post-cold war period (Feder, 2002). Towards the end of the 1980s and early 1990s, however, systematic forecasting based on statistical and computational methods started making a comeback through Schrodt's (1988, 1991) attempts to use artificial intelligence and neural networks for making predictions of behavior in international politics and interstate conflicts. While these attempts of using artificial intelligence for conflict prediction were ground breaking, they also turned out to be computationally unfeasible for any large-scale predictive modeling.[2] A decade later Zeng (1999) and Beck, King, and Zeng (2000) again highlighted the potential utility of using neural nets over logistic regression models for making predictions for conflicts, but forecasting using machine learning and artificial intelligence methods remained on the fringes of peace research.

The mid 1990s also saw the creation of State Failure Task Force (later renamed the Political Instability Task Force, PITF) (Esty et al., 1995), backed by the US government. The PITF is perhaps the first large scale modern fore-

[2]Schrodt's (1991) conflict prediction model with 310 cases and 47 predictor features took 24 hours to run. The current training data for the ViEWS (Hegre et al., 2019) prediction models on the country-month level consist of over 60,000 cases, and the prio-grid month level over 4.5 million cases.

casting and early-warning system, characterized by rigorously documented data and methodology, and continually updated forecasts (Gurr et al., 1999; Goldstone et al., 2000; Goldstone et al., 2010). While some methodological criticisms were leveled against the PITF project (King and Zeng, 2001), the PITF managed to achieve its objectives of creating practically useful forecasts with an impressive degree of accuracy (Goldstone et al., 2010; Hegre et al., 2017).

Bolstered by methodological and computational advances, advances in data collection, and, perhaps due to the successes of the PITF, increasing demands from practitioners that research should be able to produce early-warning systems for conflict, conflict forecasting started in the late 2000s to move from the fringes of peace research to becoming a more mainstream effort (Schneider, Gleditsch, and Carey, 2010; Schneider, Gleditsch, and Carey, 2011; Hegre et al., 2017). This development was further boosted by a growing criticism of traditional explanatory modelling based on null-hypothesis significance testing (NHST) (Ward, Greenhill, and Bakke, 2010; Schrodt, 2014). Ward, Greenhill, and Bakke (2010) in particular highlighted how poorly two of the most well-cited studies in the field of peace research fared when evaluated on their predictive performance, rather than their statistical significance. Driven by these factors a large number of new and ambitious forecasting projects for a wide range of outcomes including long-term trends in armed conflict (Hegre et al., 2013; Ward et al., 2013; Hegre et al., 2019), irregular leadership changes (Beger, Dorff, and Ward, 2014; Bell, Besaw, and Frank, 2021; Ward and Beger, 2017), the risk for genocide and massacres (Scharpf et al., 2014; Goldsmith et al., 2013), and non-violent uprisings (Chenoweth and Ulfelder, 2017) have been launched in the last decade. Among these projects is the ViEWS project (Hegre et al., 2019) in which this dissertation is located.

One of the consequences of the increased interest in forecasting among peace researchers is a greater methodological diversity among project forecasting different types of conflict events. By adopting tools from other branches of science such as machine-learning tools (Muchlinski et al., 2016), artificial intelligence (Colaresi and Mahmood, 2017; Malone, 2022), and simulation based approaches (Hegre et al., 2013; Scharpf et al., 2014) the field of conflict forecasting has moved away from the limitations of generalized linear models towards a truer understanding of what drives conflicts and how to make forecasts for these. The ViEWS 2020 prediction competition (Hegre, Vesco, and Colaresi, 2022) offers a great example of this methodological diversity as teams using widely different methodologies, from random forests, recurrent neural nets and pattern matching methods to hurdle models and Markov models competed to make the best out-of-sample forecasts for escalation and de-escalation of conflicts in Africa. In all, this increased methodological diversity holds great promises for the field of conflict forecasting, but also pose challenges in terms of extracting useful inferences of the drivers of conflict and producing forecasts which are possible to intuitively communicate to policy makers and other stakeholders. In addition, conflict data and processes within pose unique challenges for forecasting which require the methodology to be adapted to the unique conditions in peace research. To gain a better understanding of some of these challenges, the next section explores some differences between predictive modelling and explanatory modelling, and how predictive methods can be used to yield more useful inferences.

Prediction versus explanation

While predictive modelling and explanatory modelling are not by any means mutually exclusive, there are two important dimensions in which they differ. First, there is a theoretical difference between the scientific goals of explanation and prediction, and secondly there are practical differences between how explanatory modelling and predictive modelling is conducted and what metrics are used to evaluate models (Shmueli, 2010).

On the theoretical level, explanatory modelling can be defined as the practice of using theory to draw inferences about causal hypotheses. Explanatory modelling can be further subdivided into modelling based on observational data and explicit causal inference modelling. In observational explanatory modelling, theory is evaluated by testing causal hypotheses using association based statistical techniques such as regression on some observed, retrospective, data. In this type of modelling, the causal mechanisms are identified from theory, as the association based statistical techniques cannot by themselves evaluate these mechanisms or capture more than co-variation between different variables.

To get at these causal mechanisms more explicitly, explanatory research may instead be based on causal inference designs where the researcher either controls the identified causal mechanism directly (experiments), identify situations which have created experiment-like situations but which are outside the control of the researcher (natural-experiments), or identify variables which are known to be uncorrelated with the error in the model (instrumental variables). Such causal inference based designs are better designed to capture true causation. However, they are also more limited as it is often not possible to conduct proper experiments (or it may be unethical to do so) and it may be difficult to identify either natural experiments or potential instrumental variables. On top of these issues, explanatory modelling, whether based on causal inference techniques or observational data, is always dependent on the operationalization of the concepts which it is trying to measure, which may affect the results of the hypothesis tests. The goal of explanatory modeling is thus to gain an increased knowledge of the phenomena we are interested in and the factors which affect these phenomena (i.e. the causal mechanisms) (Shmueli, 2010; Yarkoni and Westfall, 2017).

In predictive modeling, on the other hand, the goal is to as accurately as possible predict some yet unobserved outcome of interest based on some input values. Whether or not these input values give us any useful information about how or why they affect the outcome is less interesting than whether they increase the accuracy of the predictions. Predictive modelling is therefore less concerned with the operationalization of the theoretical concepts (as long as they are functional), the causal mechanisms, and constructs such as statistical significance. Indeed, it is not uncommon for predictive models to use data reduction techniques such as the Singular Value Decomposition or Principal Component Analysis which may render the relationship between the variables virtually uninterpretable (Shmueli, 2010; Hastie et al., 2009; Yarkoni and Westfall, 2017).

In practice, these differences between explanatory and predictive modelling can be seen in four main aspects:

1. Explanatory modeling is focused on causal mechanisms while predictive modeling is focused on (observed and predicted) outcomes.
2. In explanatory modeling the models are created from theory, while in predictive modeling the models may be created from data.
3. Explanatory modeling is retrospective, i.e. only uses historical data, while predictive modeling is prospective, i.e. trying to say something about outcomes which are not yet observed.
4. Explanatory modeling is focused on minimizing bias and maximizing the certainty with which inferences are drawn, while predictive modeling is aiming to minimize the prediction error (Shmueli, 2010; Hastie et al., 2009; Yarkoni and Westfall, 2017)

To illustrate these differences, let us imagine a study where researchers are interested in the connection between infant mortality and the occurrence of localized communal violence. A researcher working with an explanatory framework may begin by considering experimental solutions, but quickly realizing that it would neither be feasible nor ethical to run an experiment with infant mortality as the experimental setting. After considering possible cases of natural experiments, the researcher also concludes that no such cases are available. The second best option may then be to simply collect observational data on infant mortality and localized communal conflict and approach the question using theory and association-based statistical analysis. In such a framework, the researchers would begin with a theoretical link between infant mortality and communal violence. They would then use the theoretical framework to generate hypotheses of how the variables should correlate if this theory was true. The researchers would then use the gathered data and test whether or not the hypothesized relationships held. Whether or not the hypothesis is useful for discriminating among new cases of communal conflict is less relevant. With a predictive approach, on the other hand, the researchers would begin with the data and try to find the combinations among all collected data which discriminated the localities where communal conflict happens from localities where communal conflict does not happen. Whether the results are useful from a theoretical point of view is less relevant than whether they are useful to predict new cases of communal conflict.

The differences between explanatory and predictive modeling have yielded very different types of methodologies for the different types of modelling, as explanatory studies require the output and results of the analysis to be easily interpretable in order to draw conclusions about the causal mechanism, while in predictive studies it is enough to be able to say the the inputs generate reasonable predictive performance for the outcome. For this reason, explanatory modeling has traditionally been using more easily interpretable association-based statistical techniques such as regression, while predictive modeling have been able to use ever more computationally intense methods such as machine learning based techniques (Hegre et al., 2017; Shmueli, 2010; Ward, Greenhill, and Bakke, 2010).

However, while explanatory and predictive modelling are clearly different scientific ventures, they are by no means mutually exclusive. Explanatory modelling can be used to generate predictions, for instance by generating out of sample predictions from a retrospective explanatory model, and predictive modelling can be used for explanation, for instance by investigating the effect of a single covariate of interest on the predicted values from a predictive model. Similarly predictive models may be constructed based on theorized relationships of causal mechanisms, and findings from predictive models may be used to generate theory (see for instance Beck, King, and Zeng, 2000; Ward et al., 2013; Hegre et al., 2019; Schrodt, 1991; Zeng, 1999). A useful example of this is the use of Markov models and Markov states in Paper III of this dissertation, where we use the researcher labelled Markov states in order to generate inferences about why we obtain the predictions we do. Another way of combining explanatory and predictive modelling is by noting that for an explanation to be truly credible it should also be possible to generate observable predictions. Thus, predictive methods may be used in order to evaluate the theoretical claims and observable predictions, and can thereby be used to validate findings from explanatory studies.

Prediction as theory evaluation

As mentioned above, one of the drivers of the increased interest in forecasting among peace researchers in the late 2000s and early 2010s was the realization that many of the explanatory studies based on the traditional explanatory framework of null-hypothesis significance testing (NHST) had very poor predictive performance. Ward, Siverson, and Cao (2007) for instance notes that using the model developed by Oneal and Russett (1999) to predict the onset of Militarized Interstate Disputes (MIDs) fails to predict *any* of the MIDs in the period 1886-1992 when using a standard cut-off value of 0.5.[3] When adjusting the cut-off value to the unconditional mean at 0.009 many of the MIDs are correctly classified, but these correctly classified MIDs are drowned out by an almost 22–1 ratio of false positives to true positives.[4] In a similar test in Ward, Greenhill, and Bakke (2010) demonstrates the poor predictive power of the models developed by Fearon and Laitin (2003) and Collier and Hoeffler (2004), which are among the most influential studies discussing the onsets of civil wars.

There are two main problems that Ward, Siverson, and Cao (2007) and Ward, Greenhill, and Bakke (2010) highlight in their critique of the NHST. The first of these problems is the difference between significant and substantial effects, i.e. that whether or not a variable is statistically significant provide very little information on whether or not this variable is useful for prediction. Indeed Ward, Greenhill, and Bakke (2010) show that certain variables from the tested models which are not statistically significant provide a substantial contribution to the predictive power of the model. The second problem these authors highlight is the problem of overfitting, i.e. when the model due to overparameterization starts fitting on the residual variance in the data rather than the underlying relationship between the variables. When a model is over-fit, it over-performs on the data it is fitted on, but under performs on data which are outside of this sample. In essence this means that the inferences drawn from the model are not generalizable beyond the specific sample they are drawn from.

Ward, Siverson, and Cao (2007) and Ward, Greenhill, and Bakke (2010) are excellent examples of how predictive methods can be used to criticize theoretical claims by testing whether or not these claims produce observable outcomes. Predictive methods can, however, also be used as robustness tests to show that the findings and theories underpinning explanatory research truly represent mechanisms which cause the outcome of interest. In essence, it is reasonable to assume that a factor which truly causes a certain outcome would also make it possible to improve predictions for that specific outcome, and if this factor does not do so it is well warranted to question whether either this factor does not at all cause the outcome or whether the effect on the outcome is relevant from a substantive perspective.

[3] I.e. when setting the bar for forecasting a MID to a predicted likelihood of a MID above 50%

[4] Setting the cut-off to the unconditional mean is the equivalent of setting the bar for forecasting a MID such that the predicted likelihood of a MID being higher than the best random guess.

Simulation as theory evaluation in processes with temporal dependency

Another approach to theory evaluation from methods primarily developed for prediction is the use of simulation based models to test claims trend-like claims in the field. Simulation based approaches need not necessarily include any theory, i.e. it is possible to construct them simply from the distribution of the outcome without regards to any inputs. Such nihilistic simulation models may, at first glance, seem irrelevant for the study of phenomena in peace and conflict research. A deeper look, however, shows that they may be utilized as a counterweight to over-theorized models, and allows us to test whether or not the trends and phenomena we see could equally well be a product of a random data generating process. Distribution based simulation has, for instance, been used to to test the empirical merits of the theory that war on the global level is in decline (Clauset, 2018; Hegre, Randahl, and Croicu, 2021), to forecast the likelihood of large-scale massacres in the conflict in Syria (Scharpf et al., 2014), and to evaluate claims about variation in the frequency and severity of terrorist attacks(Clauset, Young, and Gleditsch, 2007).

Simulation based forecasting approaches are, however, not necessarily atheoretical. Just as comparing the expectations of theory with a simulated nihilistic data generating process, the same approach can be applied to test the viability of specific theoretical claims by letting the data generating process be influenced by covariates which may or may not change over time. This type of evaluation can be particularly appealing when aiming to make predictive claims about the far future, since the random element introduced by the simulation are easily interpreted as uncertainty bounds (see for instance Clauset, 2018; Hegre et al., 2016). If the observed reality remains within these uncertainty bounds this can be interpreted as if the simulation process is consistent with the observed reality, or at least that there is not enough evidence to suggest that it is not consistent with the observed reality.[5]

Simulation is also a useful approach when aiming to predict phenomena which are highly dependent on themselves since it is a necessary condition to know the status of the phenomena at some point in the future in order to continually predict the phenomena (Hegre, Nygård, and Ræder, 2017). Such 'dynamic' simulation can thereby be used to evaluate theories with complex dependencies which may be impossible to evaluate using traditional forms of hypothesis testing. In the first paper of this dissertation, I employ such a dynamic simulation approach in order to evaluate whether or not the waves of democratization and autocratization on the global level are likely to have originated from a stable process across time.

[5]This is similar to the logic of testing a null hypothesis, i.e. we need enough evidence to suggest that we should reject the null. In this case, we need enough evidence to suggest that the simulation process is *not* consistent with the observed reality.

Summary of papers

The sections above have introduced the current state of forecasting within peace research, outlined some important differences between explanatory modelling and predictive modelling, as well as suggested ways in which predictive models may be used in order to further the goal of explaining and understanding processes of interest to peace researchers. This section introduces the four papers of the dissertation, and highlight the unique contribution of each of these papers to the growing literature on forecasting of armed conflict and contentious politics.

The four composite papers of the dissertation all explore different methodological and theoretical issues within peace research. Paper I discusses the waves of democratization and autocratization through the lens of dynamic simulation. Paper II addresses the problem of extreme values in conflict research, while Paper III makes explicit forecasts for escalation and de-escalation of armed conflicts in Africa. Lastly, Paper IV discusses the implications of missing data on predictive performance of forecasting models for making predictions of fatalities in armed conflicts. While each paper is self-contained and address different issues in peace research, there are also some overlapping themes. Papers I and III both use simulation based methodologies, Paper I dynamic simulation, and Paper III simulated Markov chains, to approach their research questions. Papers II and III are based around modelling latent states of conflict processes and the inferences and forecasts which may be drawn from such latent states. Papers I–III also all either introduce new methodologies or adapt existing methodologies from other fields to peace research. Finally, Papers III and IV are both focused on the performance of explicit forecasts of armed conflict.

Paper I: Inexorable Force or Dying wave? Long term trends of democratization and the third wave of autocratization

The paper *"Inexorable Force or Dying wave? Long term trends of democratization and the third wave of autocratization"* explores the 'waves' hypothesis of democratization on the global level. The 'waves' hypothesis of democratization was first proposed by Huntington (1991; 1993) , who noted that transitions to and from democracy seemed to cluster in time and space. Aggregated to the global level, these transitions gave rise to a wave-like pattern of democratization and autocratization over time. Huntington originally identified three waves of democracy and two 'counter-waves' of democracy.

As Huntington proposed the theory, the world was in the middle of the 'third wave of democratization' which started in the mid 1970s with the Carnation Revolution in Portugal and which then continued as it spread through Southern Europe and South America during the 1970's and 1980's. Following the end of the cold-war, the third wave of democratization was further boosted as the formerly communist countries democratized in the 1990's and during the first decade of the 21st century. In the 2010s, however, democracy on the global level started to decline (Maerz et al., 2020), sparking a debate over whether this long third wave of democratization had ended and whether or not the world had entered the 'third wave of autocratization' (see for instance Lührmann and Lindberg, 2019; Tomini, 2021; Skaaning, 2020).

However, while the empirical pattern for democratization and autocratization is clearly wave-like, there is no consensus as to why democratization and autocratization on the global level would move in such a pattern (for instance Cassani and Tomini, 2020; Skaaning, 2020; Tomini, 2021). Similarly, it is unclear whether or not the change in the global trends indicate a change in the underlying process which governs democratization and autocratization, or whether or not this wave-like pattern could appear from a stable, but highly variable, process which governs democratization and autocratization on the global level.

The paper contributes to the debate on the 'waves' hypothesis by testing whether or not the waves of democratization and autocratization in the historical record are consistent with a stable data generating process which would generate this wave-like pattern. The paper investigates this by outlining a framework through which the governance trajectory of each individual country is dependent on earlier changes in the country itself, as well as contagion from countries in the same region, and global trends. Specifically, the proposed process is anchored in critical junctures, defined as regime change, and outlines how regime changes and changes in governance tend to cluster both in space and time. This means that changes in governance is most likely to occur in the immediate aftermath of a regime change. Similarly, the likelihood of a regime change is at its greatest just following a regime change. Additionally, regime changes are contagious and tend to spread across regions, thus causing further regime changes and rapid changes in governance across regions. A wave will then randomly occur when a democratization or autocratization spreads from one country to additional countries. The wave will gather strength as further countries are swept into the wave, and will continue to spread democracy or autocracy until it either runs out of steam as the new regimes created during the wave become consolidated as the rate of democratization or autocratization decreases, or until a sufficiently large number of countries move in the opposite direction to stifle the wave and/or cause a counter-wave.

Based on this proposed process, a parsimonious simulation model is developed through which trajectories of democratization and autocratization is simulated for all independent country-days between 1789 and 2020. These simulated governance trajectories are then aggregated to the global level for each simulation and compared to the observed historical reality in order to determine whether or not the observed wave-like pattern of democratization and autocratization is consistent with the process from which the simulations are generated and whether the current trends towards autocratization is consistent with such a process.

The results show that the historical record is indeed consistent with the simulation process proposed in the paper. This means that there is not enough evidence to suggest that the underlying process which governs democratization and autocratization is not in fact a stable process which simply allows for prolonged periods (waves) of democratization and autocratization. Under this proposed process, the third wave of autocratization is also not an unlikely event as roughly one fifth of simulations see autocratization on at least the same magnitude in the early 21st century. The paper also has wider implications for the field of peace research. First, as democracy, democratization, and regime change is closely linked to the risk armed conflict and the concept of the democratic (liberal) peace (see among others Hegre, 2014; Hegre and Nygård, 2015; Dafoe, Oneal, and Russett, 2013; Oneal et al., 1996), the trends of democratization and autocratization on the global level inevitably affects our expectations for the trends of armed conflicts and the occurrence of wars. Second, the dynamic simulation system developed for this paper could feasibly be incorporated into a forecasting system for conflicts and wars on the global level. The current simulation of the paper ends at the present day, but could easily be adapted to produce forecasts for the future which could in turn inform forecasts of armed conflict (similar to Hegre et al., 2013). Additionally, the paper makes a substantive methodological contribution by showing how dynamic simulation can be used to draw inferences about the nature of processes which may be difficult to model. Similar methodology has previously been used to evaluate whether or not the 'long peace' with few large wars on the scale of the World Wars since 1945 is consistent with a sable data generating process or whether this is indicative of a change in the underlying process which governs the emergence of wars between states (Clauset, 2018; Hegre, Randahl, and Croicu, 2021).

Paper II: Inference with extremes - Accounting for Extreme Values in Count Regression Models

The second paper of the dissertation, *"Inference with extremes - Accounting for Extreme Values in Count Regression Models"*, explores methodological issues of modelling data with extreme values and proposes a new type of regression model which is able to handle data which contain both excessive zero counts and extreme values. Extreme values are a common occurrence in a wide range of scientific fields, and appear within the field of peace and conflict research when studying fort instance the number of casualties from armed conflict or one sided violence against civilians (for instance Lacina and Gleditsch, 2005; Hultman, Kathman, and Shannon, 2013), crime rates (Disha, 2019), and the number of participants in mass protests (Weidmann and Rød, 2019).

However, these processes do not always generate data which are extreme value distributed, rather they may generate data from multiple different distributions. For instance, looking at fatalities one-sided violence (OSV) against civilian on a country-month level of analysis, most country-months produce zero fatalities from OSV, while some produce a low- to middling number of fatalities from OSV, and some country-months (e.g. Rwanda in 1991) produce extremely large counts of OSV. Extreme values pose a problem for obtaining correct inferences when modelling counts using a negative binomial (NB) or zero-inflated negative binomial (ZINB) model, as the extreme values receive a very high weight in the estimation of the model. Thus, the choice of whether to include or exclude the extreme observations may have profound impact on the inferences which are drawn. Common methods for handling extreme values include trimming (excluding the extreme values) and winsorizing (setting the extreme values to the nearest non-extreme value). However, these methods are problematic in part as they alter the original distribution of values, and in part because they are largely arbitrary and researchers may therefore be vulnerable to criticism of which values are excluded/winsorized.

To handle these types of data, my co-author Johan Vegelius and I develop a new regression model, the *Extreme Value and Zero Inflated* (EVZINB) model. The inspiration for the EVZINB model stems from zero-inflated regression models where data are assumed to originate from two separate processes; one process generating zeroes, and one count process (which may also produce zeroes). In the zero-inflated regression models, these two processes are modelled separately and can be thought of as two different components of the models which may include different covariates. The first component aims to model structural, excess, zeroes using one set of covariates, and the second component models the count process separately from these 'excess zeroes'(Greene, 1994). The EVZINB model extends this framework to a three component regression model, where both excess zeroes and extreme values are modelled separately.

The benefits from this approach are manifold. First, by allowing for three separate states in the model, each of the states can be estimated while filtering out the effects of the other two. This will lead to more stable and less biased parameter estimates for each of the processes, and to fewer convergence issues. Secondly, it allows the researcher to specify different covariates on each of the processes, allowing for a more nuanced analysis. Third, filtering out the effects of extreme observations should give more stable out of sample predicted values from the model, enhancing the predictive performance of the estimated models. Fourth, the regression model allows for the estimation of the probability of any given observation being part of either of the three states, widening the possibilities for analysis further.

In the paper we showcase the EVZINB regression model by applying it to a well-known study on the effect of United Nations Peacekeeping on counts of OSV (Hultman, Kathman, and Shannon, 2013). Applying the EVZINB model to these data we show that the estimation of parameters in the EVZINB model is far more stable than the corresponding estimation in the NB and ZINB models, and that the EVZINB model outperforms both conventional modelling solutions with regards to the AIC, BIC, and cross validated out-of-sample predictive performance. Additionally, we show that the EVZINB model allows for obtaining novel and more nuanced inferences compared to the conventional NB and ZINB models.

The EVZINB model has the potential of improving the inferences which are drawn about processes which occasionally, but not always, produce extreme values in peace and conflict research. This is especially the case for studies investigating fatality counts which tend to follow a power-law distribution in the right-hand side of the distribution of values (see for instance Clauset, 2018). The EVZINB model does, however, also open up for new research questions, for instance into what causes a conflict to turn 'extreme' and at what point a conflict exceeds the threshold where it risks entering the extreme value domain. The EVZINB model could also be used in a forecasting framework to identify cases with a high risk of experiencing an extreme number of fatalities, or be incorporated into existing forecasting systems such as the ViEWS system (Hegre et al., 2019).

Paper III: Predicting escalating and de-escalating violence in Africa using Markov Models

The third paper of the dissertation, *Predicting escalating and de-escalating violence in Africa using Markov Models*, was submitted as a contribution to the ViEWS 2020 prediction competition in which a number of teams of researchers were invited to produce forecasting models for the change in the natural logarithm of battle related deaths between the time *t* and *t-s*. In the paper my coauthor, Johan Vegelius, and I develop a forecasting modelling framework based on Markov models (Jackson, 2011; Brandt et al., 2012; Diehl, 2006).

The Markov models used in this paper assume that there is an underlying state of the process which generates the change in fatalities between the two different time points. These states may either be hidden, in which case the true state is un-observed and estimated through a *hidden* Markov model, or observed, where the current and previous fatality data are used to classify the state of the model, in which case we use an observed Markov model for the predictions. The states are researcher labelled, i.e. we as reserachers define what we believe they represent, but were developed to reflect how we believed the process developed. For the observed Markov model we used four states, labelled 'peace', 'de-escalation', 'escalation', and 'conflict', and for the hidden Markov model we used two states labelled 'peace' and 'conflict'.

Transitions between the states are modelled using logistic regression, and the change in fatalities is modelled as a linear model conditional on the state. The forecasts into the future are then calculated as the weighted average of fatalities for all possible future states, where the weights are equal to the probabilities of the future states. In addition, we use a number of different covariate sets for the models which we use in a genetically weighted ensemble to produce the final predictions.

The Markov models were shown to perform well for the prediction task of the prediction competition, with the observed Markov model outperforming the benchmark model on both evaluation criteria for all time steps, and the hidden Markov model outperforming the benchmark model for all time steps on one of the evaluation criteria and performing on par for the other evaluation criteria. The Markov models also allowed inferences to be drawn about *why* the the intensity of conflicts were forecasted to change as the models also produced probabilities for the different conflict states. This allowed us to build an understanding of how the process developed over time and thereby create more intuitively understandable forecasts.

The observed Markov model developed for this paper has also proven to be a useful model for the ViEWS predicting fatalities project (Hegre et. al. 2022) where the aim is to predict the natural logarithm of fatalities on the country-month level, rather than the change in fatalities. This paper opens up for additional research into Markov processes and Markov models for use in peace research. For instance it would be entirely possible to develop a zero-inflated or extreme value and zero inflated regression model where the different components of the model is thought of as Markov states. This would especially be useful when modelling panel data with zero inflation and excess extreme values.

Paper IV: What's missing? The effect of missing data and imputation techniques on predictive performance in forecasting civil war

In the final paper of the dissertation, *What's missing? The effect of missing data and imputation techniques on predictive performance in forecasting civil war*, I explore how missing data and imputation affect forecasting models' performance. Missing data is a common problem in almost all fields of science and refers to when some observations in a data set have incomplete data. As statistical techniques generally require the analyzed data to be complete this issue has to be dealt with before the analysis is conducted by either removing observations containing missing values or by replacing the missing values with some values (imputation). It is well-known that the method for handling the missing data affects the inferences which can be drawn, as improperly handled missing data may cause parameter estimates to be biased and the associated uncertainty metrics to be either too small or too large (see for instance Lall, 2016; van Buuren, 2018; Honaker et al., 2010). However, the effects of missing data on the predictive performance of forecasting models is less studies.

In the paper I compare how different existing methods for handling missing data fare when the goal of handling the missing data is to maximize predictive performance of the models, rather than producing unbiased inferences with correct uncertainty metrics. I do this by simulating missing data in a forecasting model from the ViEWS country-month constituent models (Hegre et al., 2019) under a variety of different conditions and then using different imputation techniques to replace the missing data for the forecasts.

The results show that the advice for how to best handle missing data in inferential studies are not necessarily transferable to forecasting oriented studies. In inferential studies, the gold standard for handling missing data has long been *multiple imputation*, where missing data are imputed several times, the analysis run separately on the imputed data sets, and the results pooled. However, the results of this paper shows that simple non model based imputation strategies such as the last observation carried forward/backward and group-wise mean imputation methods, which are known to be problematic in inferential studies, as well as single imputation strategies such as the k-nearest neighbor and random forest based imputation, were the best methods for maximizing the predictive performance of the forecasting models. Among the multiple imputation strategies, only the two-level normal imputation method fared approximately on par with these single imputation methods, but at a very large computational cost. The paper also highlight under which conditions missing data is most detrimental for inferential studies. Similar to the situation in inferential studies, missing data is most problematic when data are missing at random (MAR) or missing not at random (MNAR) and when data are missing on the prediction target in the training data. This paper also offers a wider insight into the problem of missing data in prediction as it highlights how the missingness mechanism affects the predictions. Researchers should therefore base their decision on what method to use for imputing the data at least partly on what mechanism(s) they believe are responsible for generating the missing data, the underlying variability of the variables which contain missing data, and what the objective (inferential or forecasting) of the research.

Conclusions

This dissertation is dedicated to developing and improving the tools we as researchers have available to make better inferences and better predictions. The four papers of the dissertation contribute to the burgeoning field of forecasting in conflict research in a multitude of different ways. Paper I showcases how dynamic simulation may be used as a tool for evaluating theories with long temporal dependencies built into them, and also how such simulation may be used in order to make long-term forecasts for different outcomes. A key feature of this type of simulation is that it is possible to build in forecasts of several different outcomes which are interdependent. In Paper I, I connect different types of regime change with democratization and autocratization, but this may be generalized to investigate, for instance, how conflict affects different economic, social, or governance outcomes, and then in turn how the effects on the economic, social, and governance outcomes affect the risk for conflict in the future. It could be argued that in order to make reliable long-term forecasts of the risk of conflict we need to use such dynamic forecasting and simulation models in order to capture the inter-dependencies between different types of outcomes.

This use of simulation in order to make reliable forecasts is also present in Paper III, where my co-author and I employ Markov models to make forecasts of escalating and de-escalating violence. Similar to Paper I, these Markov models produce predictions into the long future based on what the model forecasts in the short future. This allows the model to create forecasts two steps into the future based on the range of potential values one step into the future. In Paper III we limit ourselves to only forecasting transition between Markov states of conflict, but as with Paper I it would be possible to extend this thinking to create interdependent Markov states for a range of different economic, social, and governance outcomes, as well as conflict outcomes in order to ask questions such as what is the effect of a (Markov) state of armed conflict, or the risk of armed conflict, on the likelihood that a country enters into an economic (Markov) state of recession, and how would a (Markov) state of recession affect the likelihood of entering into a (Markov) state of armed conflict.

Paper III also touches upon a second central theme of the dissertation, latent variables and latent states. Latent variables may in general refer to any unobserved variable which we still may want to model, but does in this dissertation primarily refer to states of processes which generate some outcome of interest. In Paper III the latent states are the either hidden or observed Markov states which govern escalation and de-escalation. In Paper II they refer to the states the process which generate one-sided violence against civilians. It is my firm belief that conflict processes in the real world are conditional on such latent states, and that finding ways of modelling this in an explicit way will help us as researchers to both better understand the processes and to make better predictions. Introducing an 'extreme value' component in the modelling of fatalities from conflict, as we do in Paper II, allows us to imagine the process as one where if we enter into this 'extreme value component' the conflict risk spiraling into much more intensive violence than what we would otherwise expect. Already in the 1960s, Lewis Fry Richardson (1960) noted that the largest conflicts seemed to follow a power-law distribution for the number of fatalities. However, not all conflicts do, so in order to model and make forecasts for large as well as small conflicts in the long-term we need to both be able to assess the likelihood of entering the more extreme process which governs the number of fatalities in large armed conflicts, as well as how smaller scale conflicts are likely to develop. While Paper II does not have an explicit aim of forecasting, this model could easily be integrated into a forecasting system for flagging conflict which may potentially spiral into very high levels of violence.

The dissertation also more broadly opens up a methodological discussion about issues which have been either neglected or under-studied in conflict research. In Paper II the main methodological puzzle is centered on the effects of extreme values on the models we use and how these extreme values can be incorporated into the analysis in an appropriate manner. In Paper IV the methodological puzzle is centered on the effect of missing data on predictive performance. While missing data at first glance may seem uninteresting to a conflict researcher, dealing with missing data is almost always part of the data pre-processing which is needed in order for the researcher to do their analysis. Without knowing how different methods for handling missing data affect the performance of models, researchers are left mostly in the dark for how to best handle these data, and may result in the use of inappropriate handling of missing data which may end up having an adverse effect on the predictive performance or inferences drawn from models. While Paper IV does not provide a definitive answer on what the new 'gold standard' of missing data imputation should be for forecasting oriented studies, it opens up for further research into these issues.

Avenues for further research

As this dissertation is primarily focused on research methodology, it opens up many avenues for further research. The tools developed and extended in this dissertation may be used for a wide range of new applications. As noted above, the dynamic simulation approach taken in Paper I may be used to further investigate and make long-term forecasts for interdependent outcomes which are difficult to make forecasts for individually. This type of modelling may also be used to approach broad questions within our field and bring clarity to questions about the long peace and other macro-level research questions our field is facing.

Paper II also opens up a wide range of new research questions which can be asked about which factors may cause a process to enter into an extreme value state where we would expect the outcome to follow a very different distribution compared to the less extreme cases. This may be used to identify conflicts or instances of one-sided violence which could turn into genocide, and may thereby guide policy makers in when and where to act to prevent conflicts from escalating beyond control. One straightforward extension of this thinking would be to combine the idea latent states with different distributions with the Markov modelling approach taken in Paper III in order to obtain both better inferences and better forecasts for when a change between latent states may occur and what would drive such a change.

The Markov modelling framework of Paper III is also generalizable to a wide range of settings. Currently this modelling approach is being used to model levels of fatalities in the ViEWS fatalities project (Hegre et al., 2022), but this approach may be used for a wide range of different outcomes. Not only does this approach allow for better predictions in many cases, it also helps us anchor the inferences we draw by using researcher labelled Markov states as a description for how the process develops over time. By using this modelling strategy we can thus both identify how certain factors or variables affect the outcome, but we can also model how the process develops over time, and what factors and variables affect this development.

Lastly, Paper IV only scratches the surface of potential research into the effects of missing data and imputation on forecasting performance. Further studies should be encourage to map out the effects of missing data and imputation on different types of estimators, and different types of data. Additionally, these studies should also investigate the effects of imputation techniques on inferences drawn on other types of models than GLMs, i.e. what are the effects of different imputation techniques on, for instance, variable importance metrics and on the estimated marginal effects of variables on the outcome.

Bibliography

Andriole, Stephen J and Robert A Young (1977). "Toward the development of an integrated crisis warning system". In: *International Studies Quarterly* 21.1, pp. 107–150.

Azar, Edward E (1980). "The conflict and peace data bank (COPDAB) project". In: *Journal of Conflict Resolution* 24.1, pp. 143–152.

Beck, Nathaniel, Gary King, and Langche Zeng (2000). "Improving quantitative studies of international conflict: A conjecture". In: *American Political science review* 94.1, pp. 21–35.

Beger, Andreas, Cassy L Dorff, and Michael D Ward (2014). "Ensemble forecasting of irregular leadership change". In: *Research & Politics* 1.3.

Bell, Curtis, Clayton Besaw, and Matthew Frank (2021). "The Rulers, Elections, and Irregular Governance (REIGN) Dataset". In: *One Earth Future*.

Brandt, Patrick T et al. (2012). "A Bayesian time series approach to the comparison of conflict dynamics". In: *APSA 2012 Annual Meeting Paper*.

Bueno de Mesquita, Bruce (1980). "An expected utility theory of international conflict". In: *American Political Science Review* 74.4, pp. 917–931.

– (1981). *The war trap*. Yale University Press.

– (1984). "Forecasting policy decisions: an expected utility approach to post-Khomeini Iran". In: *PS: Political Science & Politics* 17.2, pp. 226–236.

– (July 1988). *Forecasting political events*. New Haven, CT: Yale University Press.

Cassani, Andrea and Luca Tomini (2020). "Reversing regimes and concepts: from democratization to autocratization". In: *European Political Science* 19.2, pp. 272–287.

Chenoweth, Erica and Jay Ulfelder (2017). "Can structural conditions explain the onset of nonviolent uprisings?" In: *Journal of Conflict Resolution* 61.2, pp. 298–324.

Clauset, Aaron (2018). "Trends and fluctuations in the severity of interstate wars". In: *Science advances* 4.2, eaao3580.

Clauset, Aaron, Maxwell Young, and Kristian Skrede Gleditsch (2007). "On the frequency of severe terrorist events". In: *Journal of Conflict Resolution* 51.1, pp. 58–87.

Colaresi, Michael and Zuhaib Mahmood (2017). "Do the robot: Lessons from machine learning to improve conflict forecasting". In: *Journal of Peace Research* 54.2, pp. 193–214.

Collier, Paul (2008). *The bottom billion: Why the poorest countries are failing and what can be done about it*. Oxford University Press, USA.

Collier, Paul and Anke Hoeffler (2004). “Greed and grievance in civil war”. In: *Oxford economic papers* 56.4, pp. 563–595.
Dafoe, Allan, John R Oneal, and Bruce Russett (2013). “The democratic peace: Weighing the evidence and cautious inference”. In: *International Studies Quarterly* 57.1, pp. 201–214.
Diehl, Paul F (2006). “Just a phase?: Integrating conflict dynamics over time”. In: *Conflict Management and Peace Science* 23.3, pp. 199–210.
Disha, Ilir (2019). “Different paths: The role of immigrant assimilation on neighborhood crime”. In: *Social Science Quarterly* 100.4, pp. 1129–1153.
Esty, Daniel C et al. (1995). *State failure task force report*. McLean, VA: Science Applications International Corporation.
Fearon, James D and David D Laitin (2003). “Ethnicity, insurgency, and civil war”. In: *American political science review* 97.1, pp. 75–90.
Feder, Stanley A (2002). “Forecasting for policy making in the post–cold war period”. In: *Annual Review of Political Science* 5.1, pp. 111–125.
Gleditsch, Kristian Skrede and Michael D Ward (2013). “Forecasting is difficult, especially about the future: Using contentious issues to forecast interstate disputes”. In: *Journal of Peace Research* 50.1, pp. 17–31.
Goldsmith, Benjamin E et al. (2013). “Forecasting the onset of genocide and politicide: Annual out-of-sample forecasts on a global dataset, 1988–2003”. In: *Journal of Peace Research* 50.4, pp. 437–452.
Goldstone, Jack A et al. (2000). “State failure task force report: Phase III findings”. In: *McLean, VA: Science Applications International Corporation* 30.
Goldstone, Jack A et al. (2010). “A global model for forecasting political instability”. In: *American Journal of Political Science* 54.1, pp. 190–208.
Greene, William H (1994). *Accounting for excess zeros and sample selection in Poisson and negative binomial regression models*. NYU working paper no. EC-94-10.
Gurr, Ted Robert et al. (1999). “State failure task force report: Phase II findings”. In: *Environmental Change & Security Project Report* 5, p. 50.
Hastie, Trevor et al. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer.
Hegre, Håvard (2014). “Democracy and armed conflict”. In: *Journal of Peace Research* 51.2, pp. 159–172.
Hegre, Håvard and Håvard Mokleiv Nygård (2015). “Governance and conflict relapse”. In: *Journal of Conflict Resolution* 59.6, pp. 984–1016.
Hegre, Håvard, Håvard Mokleiv Nygård, and Ranveig Flaten Ræder (2017). “Evaluating the scope and intensity of the conflict trap: A dynamic simulation approach”. In: *Journal of Peace Research* 54.2, pp. 243–261.
Hegre, Håvard et al. (2013). “Predicting armed conflict, 2010–2050”. In: *International Studies Quarterly* 57.2, pp. 250–270.
Hegre, Håvard et al. (2016). “Forecasting civil conflict along the shared socioeconomic pathways”. In: *Environmental Research Letters* 11.5, p. 054002.

Hegre, Håvard et al. (2017). "Introduction: Forecasting in peace research". In: *Journal of Peace Research* 54.2, pp. 113–124.
Hegre, Håvard et al. (2019). "ViEWS: A political violence early-warning system". In: *Journal of peace research* 56.2, pp. 155–174.
Hegre, Håvard et al. (2021). "ViEWS2020: revising and evaluating the ViEWS political violence early-warning system". In: *Journal of peace research* 58.3, pp. 599–611.
Hegre, Håvard, David Randahl, and Mihai Croicu (2021). *Trends in the severity of wars: a mixed-level analysis*. Paper presented at the International Studies Association (ISA) Annual Convention, Las Vegas, USA.
Hegre, Håvard and The ViEWS Team (2022). *Forecasting fatalities*. Typescript Uppsala University.
Hegre, Håvard, Paola Vesco, and Michael Colaresi (2022). "Lessons from an Escalation Prediction Competition". In: *International interactions* 48.4.
Honaker, James et al. (Apr. 2010). "What to Do about Missing Values in Time-Series Cross-Section Data". In: *American Journal of Political Science* 54.2, pp. 561–581.
Hultman, Lisa, Jacob Kathman, and Megan Shannon (2013). "United Nations peacekeeping and civilian protection in civil war". In: *American Journal of Political Science* 57.4, pp. 875–891.
Huntington, Samuel P (1991). "Democracy's third wave". In: *Journal of democracy* 2.2, pp. 12–34.
– (1993). *The third wave: Democratization in the late twentieth century*. Vol. 4. University of Oklahoma press.
Jackson, Christopher (2011). "Multi-state models for panel data: the msm package for R". In: *Journal of statistical software* 38, pp. 1–28.
Kennedy, Ryan (2015). "Making useful conflict predictions". In: *Journal of Peace Research* 52.5, pp. 649–664.
King, Gary and Langche Zeng (2001). "Improving forecasts of state failure". In: *World Politics* 53.4, pp. 623–658.
Lacina, Bethany and Nils Petter Gleditsch (2005). "Monitoring trends in global combat: A new dataset of battle deaths". In: *European Journal of Population/Revue Européenne de Démographie* 21.2, pp. 145–166.
Lall, Ranjit (2016). "How multiple imputation makes a difference". In: *Political Analysis* 24.4, pp. 414–433.
Lührmann, Anna and Staffan I Lindberg (2019). "A third wave of autocratization is here: what is new about it?" In: *Democratization* 26.7, pp. 1095–1113.
Maerz, Seraphine F et al. (2020). "State of the world 2019: autocratization surges–resistance grows". In: *Democratization* 27.6, pp. 909–927.
Malone, Iris (2022). "Recurrent Neural Networks for Conflict Forecasting". In: *International Interactions* 48.4.

Moses, L. E. (1961). "Discussions and Reviews : A review: Lewis F. Richardson, Arms and Insecurity and Statistics of Deadly Quarrels". In: *Journal of Conflict Resolution vol. 5 iss. 4* 5 (4).

Muchlinski, David et al. (Dec. 2016). "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data". In: *Political Analysis* 24.1, pp. 87–103.

Oneal, John R and Bruce Russett (1999). "Assessing the liberal peace with alternative specifications: Trade still reduces conflict". In: *Journal of Peace Research* 36.4, pp. 423–442.

Oneal, John R et al. (1996). "The liberal peace: Interdependence, democracy, and international conflict, 1950-85". In: *Journal of Peace Research* 33.1, pp. 11–28.

Richardson, Lewis F (1960a). *Arms and insecurity: A mathematical study of the causes and origins of war*. Boxwood Press.

– (1960b). *Statistics of deadly quarrels*. Boxwood Press.

Scharpf, Adam et al. (2014). "Forecasting the risk of extreme massacres in Syria". In: *European Review of International Studies* 1.2, pp. 50–68.

Schneider, Gerald, Nils Petter Gleditsch, and Sabine Carey (2011). "Forecasting in international relations: One quest, three approaches". In: *Conflict Management and Peace Science* 28.1, pp. 5–14.

Schneider, Gerald, Nils Petter Gleditsch, and Sabine C Carey (2010). "Exploring the past, anticipating the future: A symposium". In: *International Studies Review* 12.1, pp. 1–7.

Schrodt, Philip A (1988). "Artificial intelligence and the study of international politics". In: *American Sociologist* 19.1, pp. 71–85.

– (1991). "Prediction of interstate conflict outcomes using a neural network". In: *Social Science Computer Review* 9.3, pp. 359–380.

– (2014). "Seven deadly sins of contemporary quantitative political analysis". In: *Journal of peace research* 51.2, pp. 287–300.

Schrodt, Philip A, James Yonamine, and Benjamin E Bagozzi (2013). "Data-based computational approaches to forecasting political violence". In: *Handbook of computational approaches to counterterrorism*. Ed. by V.S Subrahmanian. Springer, pp. 129–162.

Shmueli, Galit (Aug. 2010). "To explain or to predict?" In: *Statistical Science* 25.3, pp. 289–310.

Singer, J David (1972). "The "Correlates of War" project: Interim report and rationale". In: *World Politics* 24.2, pp. 243–270.

– (2012). "The peace researcher and foreign policy prediction (1973)". In: *Advancing Peace Research*. Ed. by Jody B Lear, Diane Macaulay, and Meredith Reid Sarkees. Routledge, pp. 102–112.

Singer, Joel David and Melvin Small (1982). *Resort to Arms: International and Civil Wars, 1816-1980*. Sage.

Singer, Joel David and Michael David Wallace (1979). *To augur well: Early warning indicators in world politics*. SAGE Publications, Incorporated.

Skaaning, Svend-Erik (2020). “Waves of autocratization and democratization: a critical note on conceptualization and measurement”. In: *Democratization* 27.8, pp. 1533–1542.
Stevens, Jacqueline (2012). “Political scientists are lousy forecasters”. In: *The New York Times*.
Tomini, Luca (2021). “Don’t think of a wave! A research note about the current autocratization debate”. In: *Democratization* 28.6, pp. 1191–1201.
van Buuren, Stef (2018). *Flexible Imputation of Missing Data, Second Edition*.
Ward, Michael D and Andreas Beger (2017). “Lessons from near real-time forecasting of irregular leadership changes”. In: *Journal of Peace Research* 54.2, pp. 141–156.
Ward, Michael D, Brian D Greenhill, and Kristin M Bakke (2010). “The perils of policy by p-value: Predicting civil conflicts”. In: *Journal of peace research* 47.4, pp. 363–375.
Ward, Michael D, Randolph M Siverson, and Xun Cao (2007). “Disputes, democracies, and dependencies: A reexamination of the Kantian peace”. In: *American Journal of Political Science* 51.3, pp. 583–601.
Ward, Michael D et al. (2013). “Learning from the past and stepping into the future: Toward a new generation of conflict prediction”. In: *International Studies Review* 15.4, pp. 473–490.
Weidmann, Nils B and Espen Geelmuyden Rød (2019). *The Internet and political protest in autocracies*. Oxford Studies in Digital Poli.
Wright, Quincy (1964). *A study of war (Second edition)*. Ed. by Louise Leonard Wright. University of Chicago Press.
Yarkoni, Tal and Jacob Westfall (2017). “Choosing prediction over explanation in psychology: Lessons from machine learning”. In: *Perspectives on Psychological Science* 12.6, pp. 1100–1122.
Zeng, Langche (1999). “Prediction and classification with neural network models”. In: *Sociological methods & research* 27.4, pp. 499–524.